

# Sample size

Janet Wittes  
Statistics Collaborative

# Sample Size

- Components of the calculation
- What the applicant should say
- What the reviewer should describe

# Biofeedback for Pain Management

- Endpoint: 10 cm VAS
  - Mean = 5
  - SD = 1

# The raw ingredients

- What is the question, precisely?
- What is the outcome, precisely?
- Who will be measured and when?
- Variability
- Handling of missing values
- Other complications (e.g., multiplicity)
- Type 1 and Type 2 error rates

# Difference to be detected

- “True” difference
  - Clinically important?
  - Biologically credible?
- Observed “I would kick myself” difference
- Affordable difference (\$)

# Does time play a role?

- Pattern of recruitment
- Follow-up time
- Hazard over time
- Hazard *ratio* over time
- Competing risks

# Operating characteristics

- Type 1 error rate = 0.05 two sided  
– (Or 0.025 one-sided)
- Type 2 error rate = 0.90

# What is the question, precisely?

- Does biofeedback control pain?
- Does does biofeedback change the mean level of pain on the VAS scale?

# What is the...

- Mean difference in score?
  - Variability:  $SD=1$
- Difference in proportions above 7?
  - Variability is binomial
- “Difference” in time to scoring 7?
  - Variability: hazard and hazard ratio

# Generic Formula

Sample size *per group*:

$$2\sigma^2 (z_1 + z_2)^2$$

---

$$\delta^2$$

$$(z_1 + z_2)^2$$

$$\frac{2\sigma^2 (z_1 + z_2)^2}{\delta^2}$$

- Subscripts usually:  $(1-\alpha)/2$  and  $(1-\beta)$
- A fixed number
- You choose!
- If 0.05 and 0.90, this quantity is about 10
- Kick-yourself power ( $\beta=.5$  and  $z_2=0$ ): value =4

$$\delta^2$$

$$\frac{2\sigma^2 (z_1 + z_2)^2}{\delta^2}$$

The difference you \_\_\_\_\_ detect.

- a) want to
- b) believe is clinically meaningful
- c) believe is biologically credible
- d) can afford to

$$\sigma^2$$

- From:
    - Past data
    - Assumptions in study
  - Very often underestimated!
    - Past data not directly relevant
    - Problems in study inflate the variance
-

2

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

- The 2 is *per group*
- The factor for a two-group study is 4

# The ideal:

## Recruitment and follow-up

- Everyone is recruited at the same time
- No one dies or is lost to follow-up
- Everyone is followed for exactly 1 year

# Endpoint: difference in mean

$$\frac{2\sigma^2 (z_1+z_2)^2}{\delta^2}$$

- Assume the mean is normal:  $\sigma=1$ ;  $\delta=1$
- Sample size =  $2 (1)(10)/1 = 20/\text{group}$
- Doubling the SD or halving the difference quadruples the sample size

# Endpoint: proportion falling above 7

- (Proportion increasing 2 points)
- (Proportion increasing at least 20 percent)
- Say we want to compare 50 percent vs. 30 percent:

$$\frac{2(\text{binomial variance})^2 (z_1 + z_2)^2}{\delta^2}$$

# Binomial answer

- 130 per group
- If only 80 percent power, 100 per group
- If Type 1 error rate is 0.01 and power =
  - 90%    n per group = 185
  - 80%    n per group = 150

# Time to hitting 7

- Assume exponential time to failure
- Assume that at 4 months 50% of control and 70% of treatment are still below 7
- Required sample size is 128 per group.

# Minor headaches

- Distribution of the mean not normal
- Population heterogeneous
- Multiple primaries
- Interim analyses

# Major headaches

- Missing data
  - Problem explicit in time-to-failure)
- Non-exponential failure
- Non-proportional hazards

# Missing data

- Common approaches
  - Just Ignore
  - Last Observation Carried Forward (LOCF)
  - Something more complicated
- My principle: you should not win because of missing data

# Noncompliance: Implications for Sample Size

- You need 100/group; expect 10% missing
  - LOCF: 100
  - Just Ignore: 111
  - Lavori: 1 noncompliant = 3 observed
    - Therefore, you need  $90 + 3(10) = 120$

# Time:

## Exponential/non-exponential

- Light bulb model often works well
- All we need to know is person-years of follow-up
- So, 4 people followed 1 year =  
1 person followed 4 years

# Recruitment: exponential case

- Follow-each person 12 months-  
recruitment pattern doesn't matter for  
sample size
- Follow each person until the last recruited  
has 12 months of follow-up
  - Persons years of follow-up depends on  
recruitment pattern
  - The SLOWER the recruitment, the  
SMALLER the sample size

# Non-exponential examples

- Post-CABG surgery:
  - Cognition impaired at first, perhaps as consequence of anesthesia
  - Long-term may show slight decline, perhaps consequence of mini-strokes

# Non-proportional hazards

- Landmark vs. log-rank time to failure
- E.g., time to diabetes
  - Control
  - Diet
  - Drug
- If we stop at two years, we have no data for four years

# What should the applicant say?

- Describe assumptions in detail
- Describe expected noncompliance
- Talk about handling those without endpoint

# The reviewer

- Should have checked:
  - assumptions
  - calculations
- Have answers similar to those in application

# Final check

- Look for tell-tale signs that the calculation was done at the last minute
- Make sure you understand  $\delta$ 's justification